Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Wiegand KC, Shah SP, Al-Agha OM, et al. *ARID1A* mutations in endometriosis-associated ovarian carcinomas. N Engl J Med 2010;363:1532-43. DOI: 10.1056/NEJMoa1008433.

Supplementary Appendix

Contents

I. Supplemental Methods and Results

Patient Samples

The 18 tumor specimens analyzed in this study for RNA sequencing (discovery cohort) were collected via a BCCA prospective tumor banking protocols for ovarian tumors (Formation of Gynaecological Tumor Bank). Patients were approached for written informed consent, before undergoing surgery, to donate tissue surplus to diagnostic requirements plus a blood sample, for use in a research ethics board (REB) approved research protocol. All patients were informed at the time of consent about the potential risks of loss of confidentiality arising from use of their samples in research and that none of the research study data would ever form part of the clinical record or be reported back to the care physicians. Specimen materials were released to investigators only with an REB approved study certificate, via a privacy guardian who anonymizes samples such that research investigators have no access to patient identifiers or information that could be used to directly identify a subject. The REB ethics protocol for genome-scale datasets stipulates that primary datasets from the tumor transcriptomes will not be released into the public domain, but can be made available via a tiered access mechanism to named investigators of institutions agreeing by a materials transfer agreement to honor the same ethical and privacy principles as the BCCA investigators.

Chart Review for Presence of Endometriosis

From out cohort of 119 CCCs (both discovery cohort and mutation validation cohort) and 33 ECs (mutation validation cohort), 86 CCCs and all 33 ECs were examined to determine if endometriosis was present at the time of surgery. These results are shown in Supplemental Table 5.

DNA and RNA Extraction from Patient Samples

DNA and RNA were extracted using standard methodologies, as previously described[1]. In cases for which insufficient DNA for *ARID1A* resequencing was available whole genome amplification (WGA) was used to extend the DNA template, however mutations were all confirmed using non-WGA treated DNA.

Paired-End RNA Sequencing and analysis (RNA-Sequencing)

RNA sequencing was performed as previously described [1]. Double stranded cDNA was synthesized from polyadenylated RNA, and the resulting cDNA was sheared. The 190-210bp DNA fraction was isolated and PCR amplified to generate the sequencing library, as per the Illumina Genome Analyzer paired end library protocol (Illumina Inc., Hayward, CA). The resulting libraries were sequenced on an Illumina GA_{ii} . Short read sequences obtained from the Illumina GA_{ii} were mapped to the reference human genome (NBCI build 36.1, hg18) plus a database of known exon junctions [2] using MAQ [3] in paired end mode. The data from RNA sequencing is summarized in Supplemental Table 1.

Single nucleotide variants (SNVs) were predicted using a Bayesian mixture model, *SNVmix*, as previously described [1, 4]. Only bases with > Q20 base quality were

considered to minimize errors. SNVs were cross-referenced against dbSNP version 129 and published genomes in order to eliminate any previously described germline variants[1]. In addition to variants in *ARID1A, CTNNB1* (C110G (S37C), NM_001904.3) somatic mutations were detected in CCC02 and CCC03 and validated by PCR amplification and Sanger sequencing in both tumor and germline DNA from these cases. Additionally, two variants were predicted based on RNA sequencing data in the TOV21G cell line in *PIK3CA* (C3139T (H1047Y), NM_006218.2) and *KRAS* (G37T (G13C), NM_004985.3) which were validated by PCR amplification and Sanger sequencing. Though variants in *BRAF* were observed in the RNA sequencing data, none of these passed validation by Sanger sequencing in tumor DNA.

RNA-Sequencing fusion prediction

Gene fusions were predicted using deFuse (manuscript in preparation). deFuse predicts gene fusions by searching paired end RNA-sequencing data for reads that harbor fusion boundaries. Spanning reads harbor a fusion boundary in the unsequenced region in the middle of the read, whereas split reads harbor a fusion boundary in the sequence of one end. deFuse searches for spanning reads with reads ends that align to different genes. Approximate fusion boundaries implied by spanning reads are then resolved to nucleotide level using dynamic programming based alignment of candidate split reads.

Copy number analysis of Affymetrix SNP 6.0 arrays

The Affymetrix SNP 6.0 arrays were normalized using CRMAv2 [5] using the default settings for performing allelic-crosstalk calibration, probe sequence effects normalization, probe-level summarization, and PCR fragment length normalization. Log ratios were then computed by normalizing against a reference generated using a normal dataset of 270 HapMap samples obtained from Affymetrix. Segmentation is performed

using an 11-state hidden Markov model (manuscript in preparation). This approach simultaneously detects and discriminates somatic and germline DNA copy number changes in cancer genomes. The hidden Markov model performs segmentation of the log ratio intensity data and predicts discrete copy number status for each resulting segment from the set of five somatic states (homozygous deletion, hemizygous deletion, gain, amplification, and high-level amplifcation), five analogous germline states, and neutral copy number. The boundaries of the segments provide candidate breakpoints in the genome as a result of copy number alteration events.

In all cases with Affymetrix SNP 6.0 data, only CCC04 contained a breakpoint in *ARID1A*. The segment (chr1:26898389-27000523) is a homozygous deletion that breaks the gene near the 5' end and truncates it. The published CNV map from 450 HapMap individuals [6] was studied to see whether any regions overlapping *ARID1A* were reported and none were found. Based on this, we predict that this is a somatic change.

Illumina-based targeted exon resequencing of ARID1A

Genomic DNA for the cases from both the discovery and mutation validation cohorts were subjected to Illumina based targeted exon resequencing. All *ARID1A* exons were PCR amplified and individual amplicons were indexed, pooled, and sequenced. Automated primer design was performed using Primer3 [7] and custom scripting. Primers were designed to span annotated exons of *ARID1A* (UCSC build hg18) with an average PCR product size of 2067bp. Primers were synthesized by Integrated DNA Technologies at a 25nmol scale with standard desalting (IDT Coralville, IA) and tested in PCR using control human genomic DNA. Primer pairs that failed to generate a product of the expected size were redesigned. The sequences for the primers are provided in Supplemental Table 2. Polymerase cycling reactions were set up in 96-well plates and comprised of 0.5 μM forward primer, 0.5 μM reverse primer, 1ng of gDNA template or

1ng of gDNA that was whole genome amplified using the REPLI-g® Mini/Midi (QIAGEN, Valencia, CA), 5X Phusion HF Buffer, 0.2 µM dNTPs, 3% DMSO, and 0.4 units of Phusion DNA polymerase (NEB, Ipswich, MA, USA). Reaction plates were cycled on a MJR Peltier Thermocycler (model PTC-225) with cycling conditions of a denaturation step at 98 °C for 30 sec, followed by 35 cycles of [98°C for 10 sec, 69°C for 15 sec, 72°C for 15 sec] and a final extension step at 72°C for 10 min. PCR reactions were visualized by SybrGreen (Life Technologies, Carlsbad, CA, USA) in 1.2% agarose (SeaKem LE, Cambrex, NJ, USA) gels run for 90min at 170V to assess PCR success. Reactions were pooled (4ul per well) by template and sheared to an average size of 200bp using a Covaris E210 ultrasonic 96 well sonication platform (75 seconds, duty cycle 20, intensity 5, cycles/burst 200; Covaris Inc. Woburn, MA) and subjected to plate based library construction on a BioMek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA) using a modified paired-end protocol (Illumina, Hayward, CA). This involved end-repair and A-tailing of sheared amplicons followed by ligation to Illumina PE adapters and PCR amplification. At each step in the process, reactions were purified using solid phase reversible immobilization paramagnetic beads (Agencourt AMPure, Beckman Coulter, Brea, CA) in 96 well plates on the BioMek FX platform using custom in house programs. Purified adapter-ligated amplicons were PCR-amplified using Phusion DNA polymerase (NEB, Ipswich, MA) in 10 cycles using PE primer 1.0 (Illumina) and a custom multiplexing PCR Primer [5' CAAGCAGAAGACGGCATACGAGATNNNNNNCGGTCTCGGCATTCCTGCTGAACCG CTCTTCCGATCT-3'] where "NNNNNN" was replaced with 96 unique fault tolerant hexamer barcodes. Individual amplicons were indexed and pooled by plate and the 200-400bp size range purified away from adapter ligation artifacts on an 8% Novex TBE PAGE gel (Invitrogen, Carlsbad, CA, USA). Individual indexes enabled the

deconvolution of reads deriving from individual samples concurrently sequenced from the same library. DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay (Agilent, Santa Clara CA) and Nanodrop 7500 spectrophotometer (Nanodrop, Wilmington, DE) and subsequently diluted to 10nM. The final concentration was confirmed using a Quant-iT dsDNA HS assay kit and Qubit fluorometer (Invitrogen, Carlsbad, CA). For sequencing, clusters were generated on the Illumina cluster station using v4 cluster reagents and paired-end 75bp reads generated using v4 sequencing reagents on the Illumina GA_{iix} platform following the manufacturer's instructions. Between the paired 75bp reads a third 7 base pair read was performed using the following custom sequencing primer [5-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG] to sequence the hexamer barcode. Image analysis, base-calling and error calibration was performed using v1.60 of Illumina's Genome analysis pipeline. The exon by exon coverage of the *ARID1A* gene from Illumina based targeted exon resequencing is shown in Supplemental Figure 2.

Data processing for ARID1A Illumina-based targeted exon resequencing

Sequence reads from the *ARID1A* targeted exon resequencing experiment were aligned to the genomic regions targeted by the PCR primers using MAQ version 0.7.1. Each exon was assessed for coverage by enumerating all uniquely aligning reads to the targeted space. SNVs were determined by computing the allelic counts for each genomic position within the complete targeted space. All positions exhibiting an allelic ratio of at least 10% variant were considered for validation by Sanger sequencing. Insertions and deletions were predicted using the Maq indelpe program using 10% allelic ratio criteria for selection for experimental follow up. In addition, to determine a confidence measure for each SNV prediction, we applied a one-tailed Binomial exact test to each position covered as described in Shah et al[1] using all aligned reads to

compute the expected distribution. Benjamini-Hochberg[8] correction for multiple comparison was applied to the resultant Binomial-test p-values to yield q-values for each position.

Sanger sequencing of ARID1A exon1

The Illumina based targeted exon sequencing of *ARID1A* did not provide coverage of exon1. To obtain sequence information for exon 1, four overlapping PCR primer sets were designed, priming sites for M13 forward and M13 reverse added to their 5′ ends to allow direct Sanger sequencing of amplicons. For the PCR, after denaturation at 94ºC for 1 min, DNA was amplified over 35 cycles (94ºC 30 sec, 58-60ºC 30sec, 72ºC 30 sec) using an MJ Research Tetrad (Ramsey, MN). Final extension was at 72ºC for 5 min. PCR products were purified using ExoSAP-IT® (USB® Products Affymetrix, Inc., Cleveland, OH) and sequenced using an ABI BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). All capillary traces were visually inspected to confirm their presence in tumor and absence from germline traces or analyzed using Mutation Surveyor.

Gene Expression Analysis

For gene expression analysis, the RNA-sequencing reads initially were mapped to the genome (NCBI36/hg 18) using MAQ (0.7.1). We used the Sequence Alignment/Map (SAMtools 0.1.7) for downstream processing. Up to five mismatches was allowed. Raw expression values (read counts) were obtained by summing the number of reads that mapped to human genes based on the Ensembl database (Release 51). The initial gene expression values were normalized using a quantile normalization procedure using

aroma.light (1.16.0.) package in R (2.11.1). Results for the 50 genes with the greatest differential expression with respect to *ARID1A* mutation status are shown in Supplemental Table 4.

Immunohistochemical (IHC) staining

Immunohistochemical (IHC) staining for BAF250a was performed in all cases included in this study with the exception of the 42 CCC from the AOCS and 4 samples from JHU. Additional IHC staining for hepatocyte nuclear factor (HNF)-1β, and estrogen receptor (ER) was performed on whole sections for two cases with associated endometriosis as previously described[9]. Immunohistochemical analysis was performed on 4μm thick paraffin sections on the semi-automated Ventana Discovery® XT instrument (Ventana Medical Systems, Tucson, AZ). ARID1A and HNF-1β was stained using the Ventana ChromoMap™ DAB kit. Antigen retrieval was standard CC1 with a two hour primary incubation. ARID1A mouse clone 3H2 (Abgent, San Diego, CA) was applied at 1:25 followed by a 16 minute secondary incubation of pre-diluted UltraMap™ Mouse HRP (Ventana). HNF-1β goat polyclonal (Santa Cruz Biotechnology, Santa Cruz, CA) was applied at 1:200 dilution followed by a 32 minute incubation of unconjugated rabbit anti-goat secondary at 1:500 (Jackson ImmunoResearch Labs Inc., West Grove, PA). Afterwards the tertiary antibody was incubated for 16 minutes with the prediluted Ventana UltraMap™ Rabbit HRP. ER immunostaining was done using the Ventana DABMap™ kit with standard CC1. The rabbit clone SP1 (Thermo Scientific, Fremont, CA) was incubated at 1:25 for 60 minutes with heat followed by a 32 minute secondary incubation with the pre-diluted Ventana Universal Secondary. Histologic images were obtained with the use of a ScanScope XT digital scanning system (Aperio Technologies Inc., Vista, CA).

Sanger sequence validation of predicted mutations

Based on the exon resequencing data, any truncating or radical missense mutations (results in change to the charge or polarity of the amino acid[10]) that occurred at an allele frequency of greater than 10% were further validated in tumor DNA, and in most cases germline DNA, using Sanger sequencing. Regions of *ARID1A* containing putative mutations were PCR amplified from genomic DNA using primers with priming sites for M13 forward and M13 reverse added to their 5′ ends to allow direct Sanger sequencing of ampliconsIn cases where the matched germline DNA of the patient was from FFPE material, short (<250nt) amplicons were designed to validate the SNVs. Unless otherwise stated, amplicons were produced from genomic DNA from both the tumor and matched germline DNA from the same patient. For the PCR, after denaturation at 94ºC for 1 min, DNA was amplified over 35 cycles (94ºC 30 sec, 60-65ºC 30sec, 72ºC 30 sec) using an MJ Research (Ramsey, MN) Tetrad. Final extension was at 72ºC for 5 min. PCR products were purified using a MinElute PCR purification kit (QIAGEN, Valencia, CA) and sequenced using an ABI BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). All capillary traces were visually inspected to confirm their presence in tumor and absence from germline traces or analyzed using Mutation Surveyor. Results from this analysis along with immunohistochemistry are summarized in Supplemental Table 3.

Laser capture microdissection (LCM), DNA isolation, and cloning

Atypical (adjacent) and distant endometriosis sections were identified in CCC13 and CCC23 by a gynecological pathologist. For microdissection, formalin-fixed paraffin-embedded (FFPE) sections (5μM) were cut on a Tissue-Tek® $Cryo_3$® cryostat (Sakura Finetek, Dublin, OH) onto clean uncharged slides. FFPE sections were deparaffanized

and rehydrated, stained with Arcturus® HistoGene® Staining Solution (Molecular Devices, Inc., Sunnyvale, CA), then dehydrated in alcohol and xylene. All reagents were prepared with nuclease-free water and all steps were performed using nuclease-free techniques.

Atypical or distant endometriotic cells were microdissected from prepared FFPE sections using the Veritas™ Laser Capture Microdissection System (Arcturus Bioscience, Inc., Mountain View, CA) according to the manufacturer's standard protocols. LCM caps with captured cells were placed directly in 15µl of lysis buffer with 10 µl of Proteinase K, and DNA was isolated using the QIAamp® DNA Micro kit (QIAGEN, Hilden, Germany). DNA was subsequently quantified on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). PCR was performed, followed by gel extraction of PCR products using the QIAquick Gel Extraction Kit (QIAGEN), PCR products were cloned using the Topo® TA Cloning® Kit following manufacturer's instructions (Invitrogen Corp., Carlsbad, CA). Inserts from individual clones were PCR amplified and Sanger sequenced to determine mutation frequency.

Fluorescent In-Situ Hybridization

Tissue samples from CCC13 and CCC23 were assayed for deletion of *ARID1A* using fluorescent *in-situ* hybridization (FISH). Six micrometer-thick sections were pre-treated as described previously [11]. Three-color FISH assays were performed using BACs specific to the regions flanking *ARID1A* (RP11-35M8 (chr1:26,609,021-26,767,926) and RP11-285H13 (chr1:27,033,759-27,216,771)) and fosmids specific to the *ARID1A* locus (G248P86703G10 (chr1:26,976,949-27,017,636), G248P89619A2 (chr1:26,954,143-26,991,761), and G248P88415D8 (chr1:26,914,023-26,954,284)). BAC and fosmid probes were obtained from British Columbia Genome Sciences Centre, and were directly labeled with Spectrum Red, Spectrum Blue, or Spectrum Green using a Nick

Translation Kit (Abbott Molecular Laboratories, Abbott Park, Il). Analysis was done on a Zeiss Axioplan epifluorescent microscope. Images were captured using Metasystems Isis FISH imaging software (MetaSystems Group, Inc. Belmont MA). Loss of heterozygosity was confirmed in CCC23 and the results were inconclusive for CCC13.

Sequencing of two mutations in CCC13

CCC13 has two somatic mutations (5541insG and T5953C (S1985P)) which were close enough to be sequenced from a single PCR fragment. PCR products were cloned and then resequenced (see Supplemental Figure 4).

II References for supplemental material

1. Shah SP, Kobel M, Senz J, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. N Engl J Med 2009;360:2719-29

2. Morin R, Bainbridge M, Fejes A, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 2008;45:81-94

3. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 2008;18:1851-8

4. Goya R, Sun MG, Morin RD, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics 2010;26:730-6

5. Bengtsson H, Ray A, Spellman P, Speed TP. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. Bioinformatics 2009;25:861-7

6. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. Nature;464:704-12

7. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132:365-86

8. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811-8

9. Kobel M, Kalloger SE, Carrick J, et al. A limited panel of immunomarkers can reliably distinguish between clear cell and high-grade serous carcinoma of the ovary. Am J Surg Pathol 2009;33:14-21

10. Dagan T, Talmor Y, Graur D. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol Biol Evol 2002;19:1022-5

11. Makretsov N, He M, Hayes M, et al. A fluorescence in situ hybridization study of ETV6-NTRK3 fusion gene in secretory breast carcinoma. Genes Chromosomes Cancer 2004;40:152-7

Supplemental Table 1. Summary of Data Generated from Whole-Transcriptome Paired-End RNA Sequencing.*

Sample number	Paired end (PE) reads	Mapped PE reads	Novel variants		Known SNPs		Genes expressed	Platform for mutation discovery
			All	Non-synonymous	All	Coding		
CCC01	64262974	48977286	3897	887	20385	8605	17326	RNA Sequencing
CCC02	81903660	68525022	5998	1095	21194	9546	17923	Exon resequencing
CCC03	65317504	36662786	10231	3197	13364	4281	16526	Exon resequencing
CCC04	79162996	70176512	4836	932	19980	9120	17200	RNA Sequencing
CCC05	81312938	41513496	9594	3309	15514	5654	17049	
CCC06	74464228	60312352	6335	1151	22661	10303	18069	RNA Sequencing
CCC09	73608312	46218944	9256	3313	15317	6037	16604	RNA Sequencing
CCC10	63192438	51692248	4744	906	17979	8275	15583	Exon resequencing
CCC13	81243638	67705474	5289	941	22592	10282	17873	RNA Sequencing
CCC14	83227842	68251298	5559	979	22855	9947	17833	RNA Sequencing
CCC66	78536592	70131816	6519	1004	21120	9335	17577	
CCC67	75424692	56887314	4729	964	20608	8352	17673	
CCC68	66162356	36977868	7609	2383	11142	3631	15934	
CCC69	77687408	62610020	6527	945	22367	9995	17896	
CCC70	108909430	73009800	12109	4145	17854	6649	17423	
CCC71	113001050	108649608	5822	792	25814	10974	18202	

*Paired-End (PE) reads: the number of two short sequences with an unsequenced insertion between the sequences resulting from sequencing of fragmented and size-selected complementary DNA from both ends.

Mapped PE reads: the number of resulting PE reads which were mapped back to the reference human genome sequence

Genes expressed: the total number of human genes expressed in each case based on the reads that mapped to each gene

RNA Sequencing: Illumina based RNA sequencing

Exon Resequencing: Illumina based targeted exon sequencing.

Supplemental Table 2: Sequence Validation Primers

ACCESSION	GENOME_REGION	PRIMER_NAME	SEQUENCE	PRIMER_NAME	SEQUENCE
ARID1A-1B	chr1:26894788+26896841	ARID1A-1BF	CTGGGTAAAATGGCAGTGCT	ARID1A-1BR	AGGAGCAGACACCTCTGGAA
ARID1A-2	chr1:26927836-26929835	SVA_008321	GCCAGTAGGATAAAGGCATATCAAAAAG	SVA_008322	ACAGTGTGAAAATCACTAAGCTACTTTGG
ARID1A-3	chr1:26929636-26931635	SVA_008323	TTTGCCTGCTTACTAAATATATCCCCAG	SVA_008324	AAGTATTATGTGAAGGCCAGAAAAATCC
ARID1A-4	chr1:26930813-26932812	SVA_008325	GCATAACCTCCTTAACTGCATAAAGACC	SVA_008326	TTTGCTACCCACCATCTTACTAGATGAC
ARID1A-5	chr1:26965973-26967972	SVA_008327	GTTGCTCTTCTCTTCCTGAACCTTACTC	SVA_008328	TCTGGTGACTACAGAACATACACAGGAG
ARID1A-6	chr1:26969251-26971350	SVA_008329	TCTAAATAAATAAGACCCTGAACACCCG	SVA_008330	CCAAAGGACCTTGTCTAACCTACTAAACC
ARID1A-7	chr1:26971101-26973100	SVA_008331	AGAAGAACATTTGAAGTTGAGGAGTGAC	SVA_008332	TACTAACTAGTTTTCAAGGCGAACCTGC
ARID1A-8	chr1:26972901-26974900	SVA_008333	CAGCAGACTACAATGTATCAACAGCAAC	SVA_008334	CTACACTAAGAAGATCCCAAACCCTCTC
ARID1A-9	chr1:26973721-26975720	SVA_008335	AAAACATGCCACCACAAATGATG	SVA_008336	GTGGGTCTGAAGAATCAGCTTTGTAAG
ARID1A-10	chr1:26979289-26981288	SVA_008337	AACTCAGCATCCAGGACAACAATG	SVA_008338	GGAGATCTCACACTAAGGGAGGTTTG
ARID1A-11	chr1:26959005-26961104	SVA_008339	CCTTATGTTGCAAAGTATGGTGACTTG	SVA_008340	TTCAGAGTGACATGAACCATTAACCAG
ARID1A-12	chr1:26960214-26962413	SVA_008341	AATATTAAGGAGCCCTAGTCTTCCACTG	SVA_008342	TAATACATTTTCTTGCACTGACACCCTC
ARID1A-13	chr1:26977525-26979924	SVA_008343	AGCTTGGACCAGTGTTTGGAGAG	SVA_008344	AAGGGCAACAGTCAGTTTCTAAGTTCTC
ARID1A-14	chr1:26963872-26966871	SVA_008345	AAATGGAAATACAACTTGACTTTGGAGG	SVA_008346	AAGTCTCTTGACTGGAGCAGTGTTAAAG

Supplemental Table 3

CASE	SOURCE	MUTATION TYPE	GENOMIC LOCATION	VARIANT	NO. OF READS CONTAINING MUTANT SEQUENCE/TOTAL NO. OF READS OVERLAPPING MUTATION	VALIDATED IN TUMOUR	VALIDATED IN GERMLINE	MUTATION STATUS	BAF250a EXPRESSION
ES2 (cell line)	VGH	missense	1:26974775	A5114G (N1705S)	1183/2652 (44.61%)	yes	not applicable	not applicable[!]	positive
TOV21G* (cell line)	VGH	indel	1:26930523	1645insC	484/1821 (26.58%)	yes	not applicable	predicted somatic[§]	negative
		indel	1:26961245	2268delC	806/2022 (39.86%)	no	not applicable	not applicable	
CCC01*[±]	VGH	indel	1:26978993	6018-6020delGCT	223//1529 (14.58%)	yes	no (FFPE)	somatic	positive
CCC02*	VGH	indel	1:26895885	404delC	exon 1	yes	no (FFPE)	somatic	negative
CCC03*[±]	VGH	indel	1:26978493	5518delG	395/1725 (22.90%)	yes	no (FFPE)	somatic	positive
CCC04*	VGH	indel	1:26898389	deletion/fusion with *ZDHHC18*[†]	not applicable	yes (by SNP 6.0)	not applicable	predicted somatic	negative
		missense	1:26928901	G1310T (R437L)	31/244 (12.70%)	no[#]	not applicable	inconclusive	
		missense	1:26930260	G1381T (G461W)	105/648 (16.20%)	no[#]	not applicable	inconclusive	
		missense	1:26930385	G1506T (Q502H)	73/518 (14.09%)	not done[#]	not applicable	inconclusive	
		missense	1:26930555	C1676A (P559H)	97/536 (18.10%)	no[#]	not applicable	inconclusive	
CCC05*[±]	VGH	missense	1:26965352	C2786A (P929H)	61/329 (18.54%)	no[#]	not applicable	inconclusive	positive
		missense	1:26979094	G6118T (G2040W)	159/1261 (12.61%)	no[#]	not applicable	inconclusive	
		missense	1:26979127	G6151T (D2051Y)	140/1392 (10.06%)	no[#]	not applicable	inconclusive	
CCC06*[±]	VGH	nonsense	1:26973506	C4201T (Q1401*)	100/914 (10.94%)	yes	no (FFPE)	somatic	positive
CCC07	VGH	indel	1:26972761	3971insA	71/528 (13.45%)	yes	no (blood)	somatic	positive
CCC08	VGH	missense	1:26970384	A3386C (K1129T)	762/2536 (30.05%)	yes	no (FFPE)	somatic	positive
CCC09*	VGH	nonsense	1:26978140	C5164T (R1722*)	1132/1513 (74.82%)	yes	no (FFPE)	somatic	negative
CCC10*	VGH	indel	1:26972738	3948delG	166/758 (21.90%)	yes	no (blood)	somatic	negative
CCC11[±]	VGH	nonsense	1:26930320	C1441T (Q481*)	204/813 (25.09%)	yes	no (FFPE)	somatic	positive
CCC12[±]	VGH	nonsense	1:26978137	C5161T (R1721*)	514/1698 (30.27%)	yes	no (blood)	somatic	positive
CCC13*[‡±]	VGH	indel	1:26978517	5541insG[¥]	395/1518 (26.02%)	yes	no (blood)	somatic	negative
		missense	1:26978929	T5953C (S1985P)	540/1712 (31.54%)	yes	no (blood)	somatic	
CCC14*[±]	VGH	nonsense	1:26930559	C1680A (Y560*)	1411/2651 (53.23%)	yes	no (blood)	somatic	negative
CCC15[±]	VGH	nonsense	1:26978542	C5566T (Q1856*)	447/1299 (34.41%)	yes	no (blood)	somatic	positive
CCC16[±]	VGH	indel	1:26895767	286-296delGCGGAGCCGGA	exon 1	yes	no (blood)	somatic	positive
CCC17[‡±]	VGH	nonsense	1:26928776	T1185A (Y395*)	392/787 (49.81%)	yes	no (blood)	somatic	negative
		indel	1:26974013	4709-4712delCTAA	426/2303 (18.50%)	yes	no (blood)	somatic	
CCC18[±]	VGH	nonsense	1:26928768	C1177T (Q393*)	51/387 (13.18%)	yes	no (blood)	somatic	negative
CCC19[±]	VGH	missense	1:26978470	G5494T (G1832W)	83/826 (10.05%)	no[#]	not applicable	inconclusive	positive
CCC20[‡±]	VGH	nonsense	1:26930245	C1366T (Q456*)	512/2238 (22.88%)	yes	no (blood)	somatic	negative
		indel	1:26978280	5305delC	950/2225 (42.70%)	yes	no (blood)	somatic	
CCC21	VGH	missense	1:26978591	C5615T (A1872V)	838/1078 (77.74%)	yes	yes (FFPE)	germline	positive
CCC22[±]	VGH	indel	1:26974657	4997delC	695/3171 (21.92%)	yes	no (FFPE)	somatic	negative
CCC23*[±]	VGH	nonsense	1:26979115	G6139T (E2047*)	952/1374 (69.29%)	yes	no (blood)	somatic	negative
CCC24	Montreal	nonsense	1:26961284	C2306G (S769*)	505/1386 (36.44%)	yes	germline DNA not available	somatic	negative
		nonsense	1:26972770	C3979T (Q1327*)	147/751 (19.57%)	no[#]	germline DNA not available	inconclusive	
		missense	1:26978525	A5549G (D1850G)	92/739 (12.45%)	no[#]	germline DNA not available	inconclusive	
CCC25	Montreal	indel	1:26978269	5294delA	670/1823 (36.75%)	yes	no (blood)	somatic	negative
CCC26	Montreal	nonsense	1:26972779	C3988T (Q1330*)	265/1570 (16.88%)	no[#]	not applicable	inconclusive	positive
CCC27	Montreal	nonsense	1:26930449	C1570T (Q524*)[¥]	310/786 (39.44%)	yes	no (blood)	somatic	positive
		missense	1:26965379	C2813T (A938V)	97/612 (15.85%)	not done[#]	no (blood)	inconclusive	
		indel	1:26979390	6415insC	779/6650 (11.71%)	no	not applicable	did not validate	
CCC28	Montreal	missense	1:26973770	G4465C (A1489P)	354/3084 (11.48%)	no[#]	not applicable	inconclusive	positive
		missense	1:26974082	C4777T (R1593W)	230/1782 (12.91%)	no[#]	not applicable	inconclusive	
CCC29	Montreal	indel	1:26978271	5296insG	613/1797 (34.11%)	yes	no (blood)	somatic	negative
		indel	1:26978517	5541insG[¥]	141/1231 (11.45%)	no	not applicable	did not validate	
CCC30	Montreal	indel	1:26962348	2718-2719insCT	2445/4619 (52.93%)	yes	no (blood)	somatic	negative
CCC31	Montreal	indel	1:26896589	1108delG	exon 1	yes	no (blood)	somatic	negative
CCC32	Montreal	indel	1:26896551	1070delG	exon 1	yes	not done[¶]	predicted somatic	negative
		indel	1:26979722	6747-6748delAG	341/3401 (10.03%)	no	not applicable	did not validate	
CCC33	Montreal	indel	1:26978976	6001-6002insCA	477/2309 (20.66%)	yes	no (blood)	somatic	negative
CCC34[‡]	Montreal	indel	1:26973944	4640delC	583/2274 (25.64%)	yes	no (blood)	somatic	negative
		indel	1:26979639	6664-6665delCC	380/2192 (17.34%)	yes	no (blood)	somatic	
CCC35[‡]	Montreal	indel	1:26930654	1776delT	470/2074 (22.66%)	yes	no (blood)	somatic	negative
		indel	1:26961268	2291insC	527/1569 (33.59%)	yes	no (blood)	somatic	

CCC36	Montreal	missense	1:26974298	G4993A (G1665R)	1177/3894 (30.23%)	PCR failure	not applicable	inconclusive	negative
		indel	1:26895496	14-41insGGTCGCCCCCCGCCTCCAGCACCTGGGC	exon 1	yes	no (blood)	somatic	
CCC37‡	Montreal	nonsense	1:26930506	C1627T (Q543*)¥	422/1262 (33.44%)	yes	no (blood)	somatic	negative
		missense	1:26979172	A6196G (N2066D)	759/2003 (37.89%)	yes	no (blood)	somatic	
CCC38	Montreal	indel	1:26978517	5541insG¥	460/1104 (41.67%)	yes	no (blood)	somatic	negative
CCC39‡	Australia	indel	1:26930276	1398delC	359/1101 (32.61%)	yes	no (blood)	somatic	not available
		indel	1:26960119	2107delC	229/1147 (19.97%)	yes	no (blood)	somatic	
		missense	1:26965568	G2912C (G971A)	64/569 (11.25%)	not done#	not applicable	inconclusive	
CCC40‡	Australia	indel	1:26895632	150insG	exon 1	yes	no (blood)	somatic	not available
		indel	1:26896589	1108insG	exon 1	yes	not done¶	predicted somatic	
		indel	1:26978517	5541insG¥	136/1173 (11.59%)	no	not applicable	did not validate	
CCC41±	Australia	missense	1:26979278	A6302G (D2101G)	379/1727 (21.95%)	no	not applicable	did not validate	not available
		nonsense	1:26979631	C6655T (Q2219*)	175/1496 (11.70%)	no#	not applicable	inconclusive	
CCC42±	Australia	indel	1:26965566	2911insG	158/722 (21.88%)	yes	no (blood)	somatic	not available
CCC43	Australia	nonsense	1:26895695	C214T (E72*)	exon 1	yes	not done¶	predicted somatic	not available
CCC44±	Australia	indel	1:26895601	120-145delGGCGGCGGCAGCGGCCGAGCGCGGGG	exon 1	yes	not done¶	predicted somatic	not available
CCC45±	Australia	nonsense	1:26930506	C1627T (Q543*)¥	659/1442 (45.70%)	yes	no (blood)	somatic	not available
CCC46‡	Australia	nonsense	1:26930521	C1642T (Q548*)	278/651 (42.70%)	yes	not done¶	predicted somatic	not available
		nonsense	1:26973887	C4582T (R1528*)	574/1408 (40.77%)	yes	not done¶	predicted somatic	
CCC47±	Australia	indel	1:26971689	3519delC	787/2196 (35.84%)	yes	no (blood)	somatic	not available
CCC48	Australia	indel	1:26930676	1798insC	306/994 (30.78%)	yes	no (blood)	somatic	not available
CCC49	Australia	nonsense	1:26930449	C1570T (Q524*)¥	515/1006 (51.19%)	yes	no (blood)	somatic	not available
CCC50‡±	Australia	indel	1:26974155	4851delT	454/1810 (25.08%)	yes	not done¶	predicted somatic	not available
		indel	1:26895538	57-62insGCCGCC (looks homozygous!)	exon 1	yes	not done¶	predicted somatic	
CCC51	Australia	indel	1:26928787	1197delC	578/2043 (28.29%)	yes	no (blood)	somatic	not available
		missense	1:26928788	C1197A (N399K)	55/292 (18.84%)	no#	not applicable	inconclusive	
		indel	1:26928790	1200delA	2020/2024 (99.80%)	no	not applicable	did not validate	
CCC52	Australia	missense	1:26978351	G5375A (G1792D)	427/3073 (13.90%)	no#	not applicable	inconclusive	not available
CCC53‡±	Australia	indel	1:26895951	470-486delGCCCGTCTGCCGTCGCC	exon 1	yes	no (blood)	somatic	not available
		nonsense	1:26973929	G4624T (E1542*)	1547/3013 (51.34%)	yes	not done¶	predicted somatic	
		missense	1:26978534	A5558G (E1853G)	176/1232 (14.29%)	no#	not applicable	inconclusive	
CCC54±	Australia	nonsense	1:26974145	C4840T (Q1614*)	1697/2306 (73.59%)	yes	no (blood)	somatic	not available
CCC55	Australia	indel	1:26973691	4387insA	841/1685 (49.91%)	yes	no (blood)	somatic	not available
CCC56±	Australia	indel	1:26973434	4130insA	347/1365 (25.42%)	yes	no (blood)	somatic	not available
		indel	1:26973435	4131insC	1352/1355 (99.78%)	no	not applicable	did not validate	
CCC57	JHU	indel	1:26896215	734-747delCGGCTGCCGGCTCC	exon 1	yes	no (FFPE)	somatic	not available
CCC58±	JHU	indel	1:26896488	1007delG	exon 1	yes	not done¶	predicted somatic	not available
CCC59	JHU	missense	1:26974775	A5114G (N1705S)	852/2166 (39.34%)	yes	yes (cultured fibroblasts)	germline	positive
CCC60	JHU	indel	1:26972006	3660-3684delGATGGGGCGCATGTCCTATGAGCCA	285/759 (37.55%)	yes	no (cultured fibroblasts)	somatic	negative
CCC61	JHU	indel	1:26972885	4011-4012delTT	1158/6957 (16.65%)	yes	no (cultured fibroblasts)	somatic	negative
CCC62±	JHU	indel	1:26974232	4928-4931insCTGG	537/3019 (17.79%)	no	not applicable	did not validate	negative
CCC63±	JHU	indel	1:26974232	4928-4931insCTGG	551/2698 (20.42%)	no	not applicable	did not validate	not available
CCC64±	JHU	indel	1:26972561	3852insA	463/1589 (29.14%)	yes	no (FFPE)	somatic	negative
CCC65‡±	JHU	indel	1:26978875	5900-5901insTG	359/1075 (33.40%)	yes	no (cultured fibroblasts)	somatic	negative
		nonsense	1:26896034	C553T (Q185*)	exon 1	yes	no (cultured fibroblasts)	somatic	
EC01	VGH	indel	1:26896510	1029-1043delAGCTGCGGCGGCGGC	exon 1	yes	not done¶	predicted somatic	positive
EC02±	VGH	missense	1:26973924	A4619G (N1540S)	285/1985 (14.36%)	no#	not applicable	inconclusive	positive
		missense	1:26973927	A4622C (H1541P)	194/1775 (10.93%)	no#	not applicable	inconclusive	
EC03±	VGH	nonsense	1:26978941	C5965T (R1989*)¥	554/1474 (37.58%)	yes	no (blood)	somatic	negative
EC04‡	VGH	nonsense	1:26930566	C1687T (Q563*)	790/2002 (39.46%)	yes	no (blood)	somatic	negative
		indel	1:26970208	3211delA	1100/2999 (36.68%)	yes	no (blood)	somatic	
EC05	VGH	nonsense	1:26962317	G2686T (E896*)	358/2075 (17.25%)	no#	not applicable	inconclusive	negative
EC06‡	VGH	indel	1:26961268	2291delC	432/2160 (20.00%)	yes	no (FFPE)	somatic	negative
		indel	1:26979390	6415delC	1820/6232 (29.20%)	yes	no (FFPE)	somatic	
EC07‡±	VGH	indel	1:26962055	2425insT	85/639 (13.30%)	yes	no (blood)	somatic	negative
		nonsense	1:26896286	C805T (Q269*)	exon 1	yes	no (blood)	somatic	
EC08	VGH	nonsense	1:26979448	C6472T (R2158*)	339/1093 (31.02%)	yes	no (blood)	somatic	positive
		missense	1:26895837	A356G (E119G)	exon 1	yes	yes (blood)	germline	
EC09	VGH	nonsense	1:26896126	C645A (Y215*)	exon 1	yes	not done¶	predicted somatic	positive
EC10±	VGH	indel	1:26978517	5539delG	151/873 (17.30%)	yes	no (FFPE)	somatic	positive
		missense	1:26965612	G2956T (D986Y)	19/170 (11.18%)	no#	not applicable	inconclusive	

EC11	VGH	missense	1:26965622	C2966A (P989H)	17/163 (10.43%)	no[#]	not applicable	inconclusive	positive
		missense	1:26973942	C4637A (P1546H)	143/1397 (10.24%)	no[#]	not applicable	inconclusive	
		missense	1:26978107	G5131T (G1711W)	91/829 (10.98%)	no[#]	not applicable	inconclusive	
		missense	1:26978753	G5777T (G1926V)	137/1087 (12.60%)	no[#]	not applicable	inconclusive	
		missense	1:26979031	C6055A (H2019N)	131/1100 (11.91%)	no[#]	not applicable	inconclusive	
EC12[‡±]	VGH	indel	1:26962057	2427delC	381/2381 (16.00%)	yes	no (FFPE)	somatic	negative
		nonsense	1:26978479	C5503T (Q1835*)	533/1740 (30.63%)	yes	no (FFPE)	somatic	
EC13[‡±]	VGH	missense	1:26973470	T4165C (Y1389H)	529/1363 (38.81%)	yes	no (blood)	somatic	positive
		nonsense	1:26978941	C5965T (R1989*)[¥]	397/958 (41.44%)	yes	no (blood)	somatic	
HGS01	VGH	missense	1:26978591	C5615T (A1872V)	310/1918 (16.16%)	yes[#]	yes (FFPE)	germline	positive
HGS02	VGH	missense	1:26960136	A2123C (Q708P)	667/1500 (44.47%)	yes	yes (FFPE)	germline	positive
HGS03	VGH	missense	1:26973501	A4196C (Q1399P)	655/1154 (56.76%)	yes	yes (FFPE)	germline	positive
HGS04	VGH	missense	1:26896062	C581T (P194L)	exon 1	yes	yes (FFPE)	germline	positive
HGS05	VGH	missense	1:26960136	A2123C (Q708P)	1145/1439 (79.57%)	yes	yes (FFPE)	germline	positive
HGS06	VGH	missense	1:26895710	G229A (A77T)	exon 1	yes	yes (blood)	germline	positive
HGS07	VGH	missense	1:26973470	T4165C (Y1389H)	170/907 (18.74%)	no[#]	not applicable	inconclusive	positive
		nonsense	1:26978941	C5965T (R1989*)[¥]	428/1569 (27.28%)	no	not applicable	did not validate	

LEGEND

* cases analysed by RNA sequencing

‡ cases with multiple (more than one) somatic or predicted somatic mutations

± cases with associated endometriosis

¥ recurrent somatic/predicted somatic mutation

mutant allele frequency less than 20%

¶ presumed somatic in absence of germline DNA analysis because mutation is truncating (see Results)

↑identical change is a germline variant in CCC59, thus not assumed to be somatic change in ES2 cells

§ cell line variant is assumed to occur as somatic mutation in CCCs

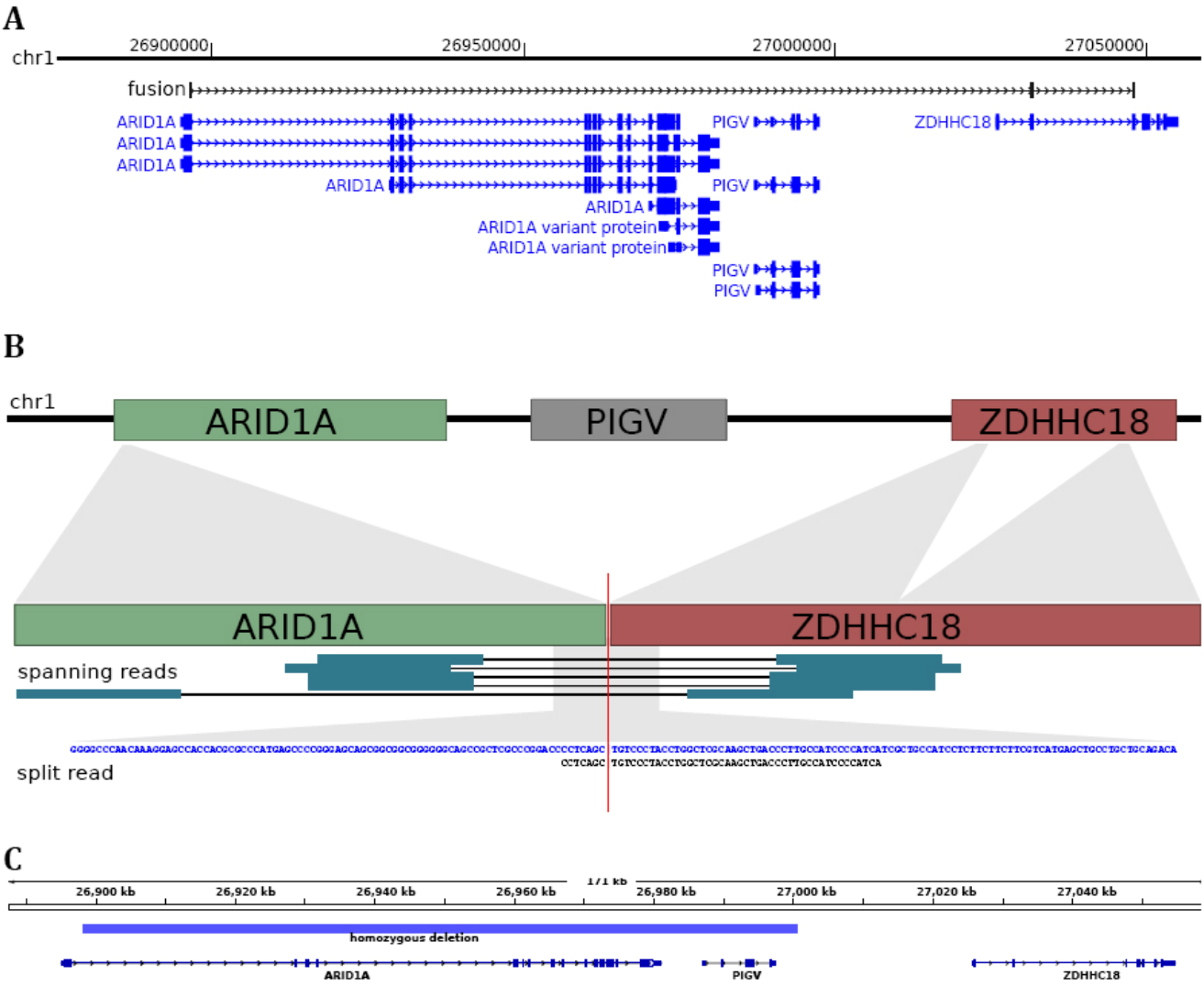
† for details of deletion/fusion see Supplemental Figure 1

Supplemental Table 4: Differential Expression of Genes based on *ARID1A* Mutation Status

Gene Symbol	Fold Change (cases with ARID1A mutations/cases with normal ARID1A)	p-value	q-value
KRT36	-5.9156	2.342E-07	1.814E-03
NKAIN1	6.5428	1.116E-04	3.751E-01
PROP1	4.0628	1.453E-04	3.751E-01
C19orf57	1.8549	6.254E-04	8.186E-01
FAM55D	-4.5453	6.831E-04	8.186E-01
ABCG8	5.2462	7.745E-04	8.186E-01
CRIM2	-1.9531	7.948E-04	8.186E-01
ZMYND12	1.8983	1.022E-03	8.186E-01
RNF183	-2.7073	1.120E-03	8.186E-01
AC138761.4	1.9836	1.216E-03	8.186E-01
GPR55	-4.8946	1.293E-03	8.186E-01
GDF3	4.6608	1.300E-03	8.186E-01
CPA6	-5.9903	1.417E-03	8.186E-01
C20orf141	-5.8876	1.504E-03	8.186E-01
NTRK1	-4.3039	1.612E-03	8.186E-01
SEMA3G	1.4987	1.698E-03	8.186E-01
LRRC2	2.6049	1.797E-03	8.186E-01
AC004522.3	3.5637	2.007E-03	8.637E-01
WDR21B	1.9457	2.290E-03	9.074E-01
SLFN12L	-2.0385	2.343E-03	9.074E-01
PAK7	4.5352	2.613E-03	9.639E-01
SPDYC	4.8633	3.205E-03	9.663E-01
THRB	2.4309	3.242E-03	9.663E-01
TEKT5	2.5372	3.458E-03	9.663E-01
PROKR1	4.2852	3.506E-03	9.663E-01
STAT4	-1.7912	3.556E-03	9.663E-01
C9orf11	-4.3519	3.839E-03	9.663E-01
CD6	-2.1419	3.950E-03	9.663E-01
AC104073.3	4.6765	3.977E-03	9.663E-01
FRAT2	1.6796	4.033E-03	9.663E-01
HSPB8	-2.4558	4.240E-03	9.663E-01
BMP5	6.0111	4.321E-03	9.663E-01
AC005826.1	4.3893	4.457E-03	9.663E-01
PXMP4	1.1051	4.849E-03	9.663E-01
IL17D	4.3644	4.942E-03	9.663E-01
TPPP3	1.6509	5.013E-03	9.663E-01
CHI3L1	-3.5315	5.190E-03	9.663E-01
PEX11G	1.7031	5.304E-03	9.663E-01
AC114341.9	4.2712	5.320E-03	9.663E-01
ALK	3.1822	5.899E-03	9.663E-01
LGI3	-3.7262	6.133E-03	9.663E-01
FCRL2	-5.2650	6.445E-03	9.663E-01
CLCA3	-2.9996	6.511E-03	9.663E-01
TRIB3	-1.6674	6.627E-03	9.663E-01
KLF17	-3.6395	6.658E-03	9.663E-01
OTUD7A	1.6996	6.749E-03	9.663E-01
KIR3DX1	-3.5428	6.812E-03	9.663E-01
DYNC1I1	1.8960	7.085E-03	9.663E-01
SIGLEC6	-3.1026	7.195E-03	9.663E-01
S100A5	-2.0973	7.626E-03	9.663E-01

Supplemental Table 5: *ARID1A* Mutation Status and Presence of Endometriosis

ARID1A Mutation Status	normal *ARID1A*						mutant *ARID1A*					
Endometriosis	endometriosis			no endometriosis			endometriosis			no endometriosis		
BAF250a IHC	positive	negative	not available	positive	negative	not available	positive	negative	not available	positive	negative	not available
CCC	11	1	10	7	2	17	7	9	9	2	3	8
EC	11	0	0	10	2	0	2	3	0	3	2	0



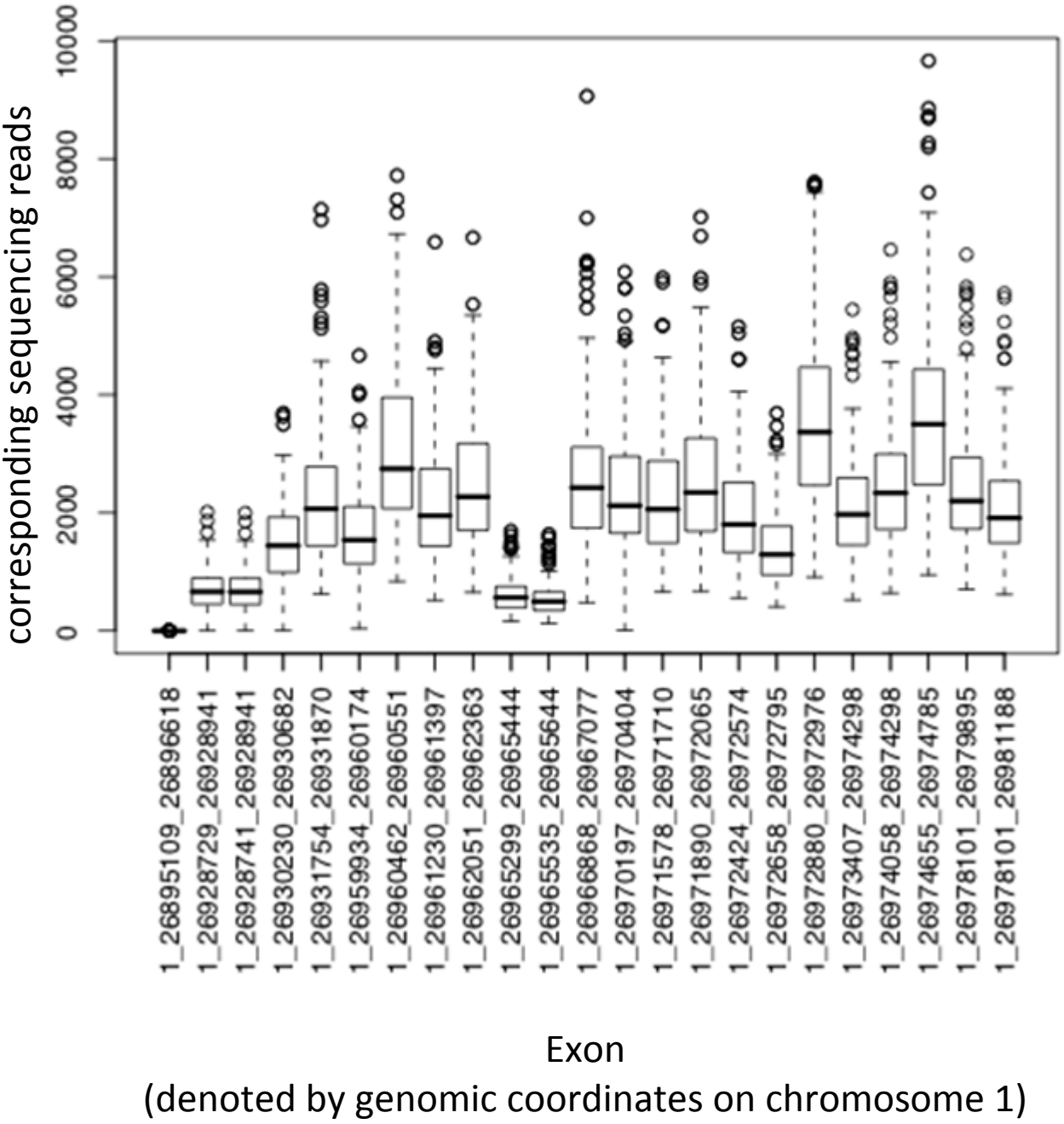
Supplemental Figure 1:

ARID1A-ZDHHC18 fusion prediction

Alignment of the predicted fusion sequence showing alignment of one end to ARID1A and the other end to ZDHHC18 at 1p36. Verticle bars in the “fusion” alignment represent exon alignment over the genomic locus.

Evidence for the ARID1A-ZDHHC18 fusion transcript includes 5 spanning reads and 1 read split by the fusion boundary.

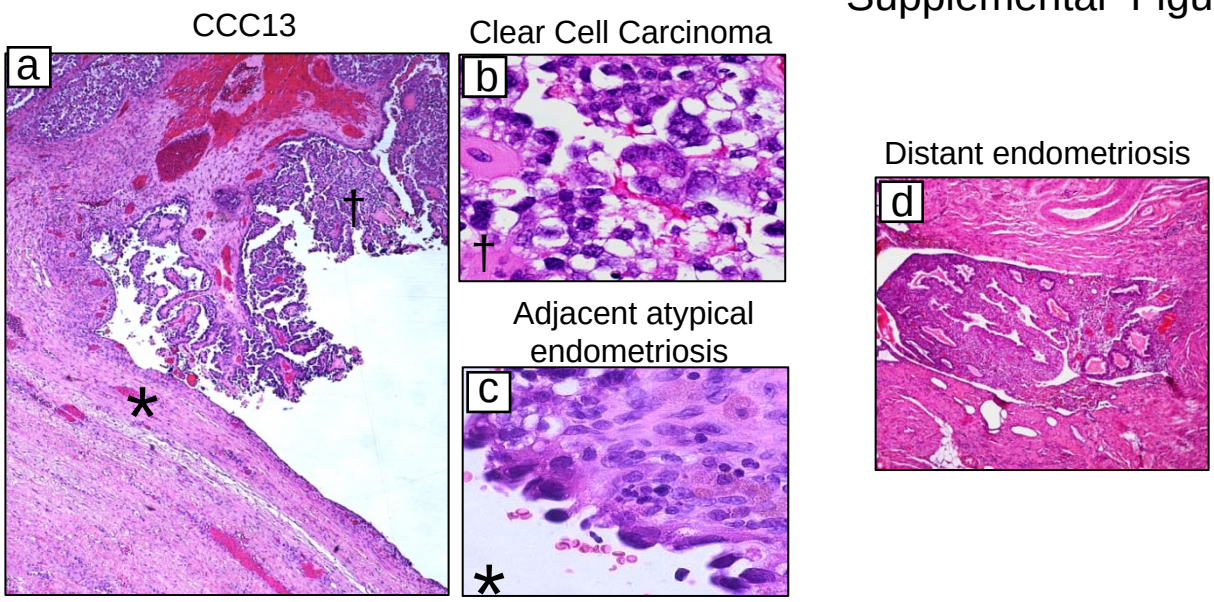
Copy number analysis from SNP 6.0 array shows a large homozygous deletion encompassing almost the entirety of ARID1A with the exception of the 5’ end, and extending over the neighboring gene PIGV.



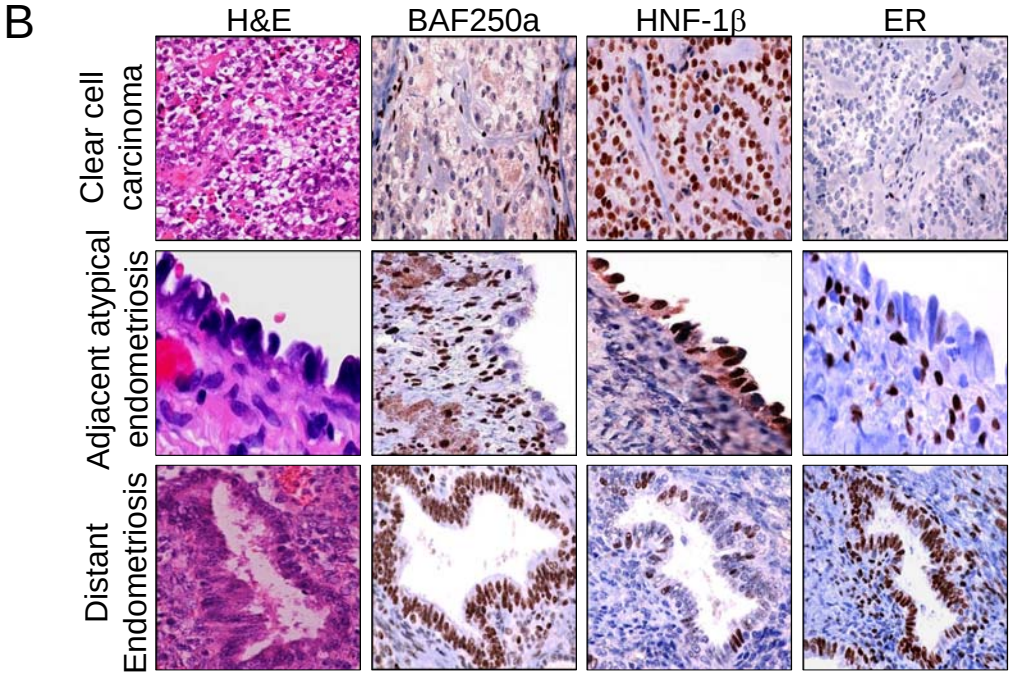
Supplemental Figure 2: Illumina based resequencing of individual ARID1A exons. Boxplots represent read coverage from each exon for every sample, open circles denote outlier samples, and the bold line represents the median coverage. Exons 1-20 are denoted by their chromosomal coordinates on chromosome 1 along the x-axis. Note that exon 1 resequencing was largely ineffective by this method and was completed by Sanger sequencing (see also Methods and Supplemental Methods).

Supplemental Figure 3

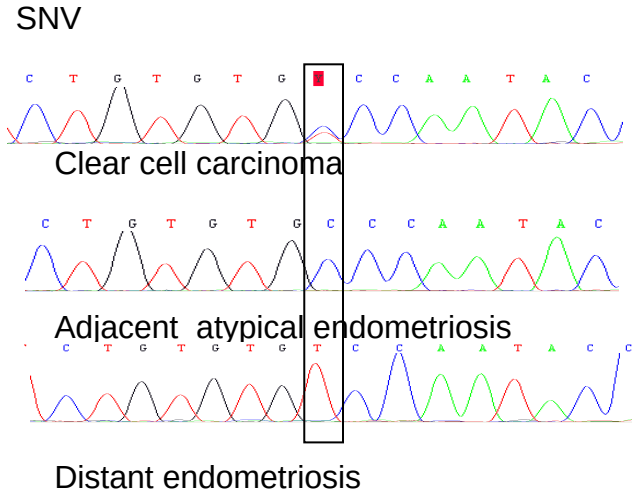
A



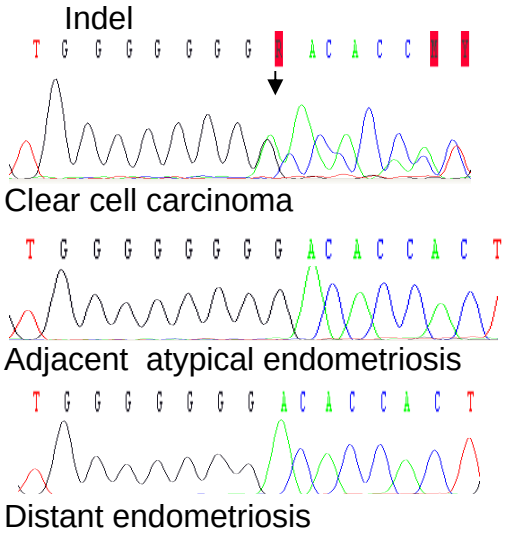
B



C



D

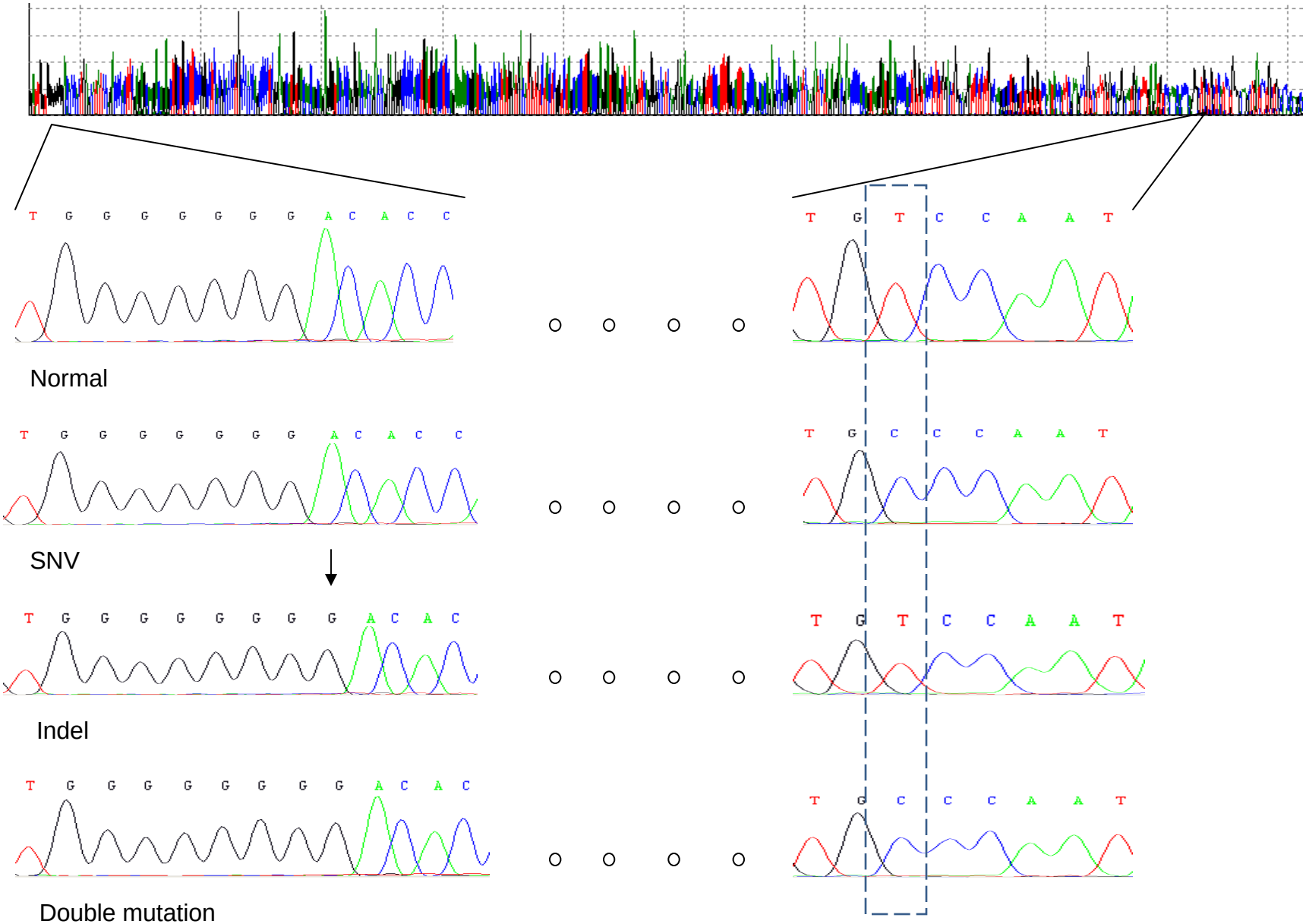


Supplemental Figure 3:

Clear cell carcinoma (specimen CCC13) and adjacent atypical endometriosis.

A. H&E stained sections from clear cell carcinoma (*) arising in an endometriotic cyst (†) at low power showing adjacent histologies (a), and at higher power showing regions of the clear cell carcinoma (b) and adjacent atypical endometriosis (c). A distant region of endometriosis from the same individual is shown at low power (d).
B. BAF250a immunoreactivity is lost in the epithelial portion of both the clear cell carcinoma and adjacent atypical endometriosis, however is maintained in the distant endometriosis. HNF-1β can be seen in both the tumor and the adjacent atypical endometriosis, however is largely negative with only occasionally positive cells in the distant endometriosis. ER is highly expressed only in the distant endometriosis and is lost in both the tumor sample and adjacent atypical endometriosis.
C. Sequencing chromatograms from the clear cell carcinoma and a PCR clone from contiguous atypical endometriosis clearly show the nucleotide variation corresponding to T5953C (S1985P). This mutation was present in 20/51 clones from the contiguous atypical endometriosis. In contrast, all cloned PCR products (from 58 clones) from distant endometriosis, which maintained BAF250a expression, show only wild type sequence. A heterozygous peak is seen in the DNA from the tumor. Micro-dissected material from both endometriosis samples was used to extract DNA, amplify by PCR, clone and sequence. None of the PCR clones from the distant endometriosis showed variation from the wild-type sequence.
D. As in “C” Sequencing chromatograms from the clear cell carcinoma and a PCR clone from contiguous atypical endometriosis show an insertion of an additional G (5541InsG). This mutation was present in 3/54 clones from the contiguous atypical endometriosis. In contrast, all cloned PCR products (from 59 clones) from the distant endometriosis, which maintained BAF250a expression, show only wild type sequence. Sequencing read from the tumor sample shows characteristic overlapping reads corresponding to the in frame and out of frame alleles after the insertion point. As in “C” sequence from PCR clones are shown for both adjacent atypical endometriosis and distant endometriosis.

Supplemental Figure 4



Supplemental Figure 4:

Sanger sequencing results from CCC13. The two somatic mutations (5541insG and T5953C (S1985P)) were sequenced from a single PCR fragment. PCR products were cloned and then resequenced. In total, sequences from 45 clones were analyzed. We found 15/45 (33%) wildtype sequence, 9/45 (20%) sequences with the T5953C (S1985P) mutation, 9/45 (20%) sequences with the 5541insG mutation, and 12/45 (27%) sequences with both mutations in a single Sanger sequence trace. This reveals the complex relationship between the mutations which occur both in trans (on independent alleles) and also in cis (on the same allele) (see Supplemental Figure 4). This finding along with the presence of wildtype alleles, suggest that this tumor is aneuploid and a gene conversion or other rearrangement at the *ARID1A* locus has occurred and is present in a subset of cells.